3-1-2008

Tutorial: Web Analytics – A Brief Tutorial

Robert J. Boncella
Washburn University, bob.boncella@washburn.edu

Web Analytics: A Brief Tutorial

Dr. Robert J. Boncella
Washburn University, School of Business
bob.boncella@washburn.edu

ABSTRACT

Web analytics is the study of the behavior of website visitors. In a commercial context, web analytics refers to the use of data collected from a web site to determine which aspects of the website achieve the business objectives. This tutorial presents the context of Web Analytics within Decision Support Systems; the technology required for Web Analytics; the data collection and data analysis techniques used in Web Analytics and an overview of a software tool used for web analytics and several case studies illustrating its use. This paper is intended for those who wish to gain an overview of the purpose of Web Analytics and what is required for Web Analytics. Additionally it is intended for those who have an in interest how Web Analytics would fit into a Market Research Techniques course.

Keywords

Business Intelligence, Business Analytics, Web Analytics, Web Mining

INTRODUCTION

Web analytics is the study of the behavior of website visitors. In a commercial context, web analytics refers to the use of data collected from a web site to determine which aspects of the website achieve the business objectives. The intent of this tutorial is to introduce those who are unfamiliar with what Web Analytics is and what is required in terms of technology, data collection and data analysis.

The tutorial presentation may be found at: www.washburn.edu/faculty/boncella/Web Analytics Tutorial.ppt .

CONTEXT FOR WEB ANALYTICS

To set the context for Web Analytics a definition of each of the following is provided: Decision Support Systems (DSS), Business Intelligence (BI), Business Analytics (BA), and Web Analytics (WA).

DSS – A conceptual framework for a process of supporting managerial decision- making, usually by modeling problems and employing quantitative models for solution analysis
BI - An umbrella term that combines architectures, tools, databases, applications, and methodologies
BA - The application of quantitative models directly to business data in order to assist in making strategic decisions
WA - The application of business analytics activities to Web-based processes, including e-commerce

The containment relationship among these is: WA is a subset of BA and BA is subset of BI and BI is subset of DSS.

TECHNOLOGY USED IN WEB ANALYTICS

The technologies used in Web Analytics are: the Internet & TCP/IP; Client / Server Computing; HTTP (Hypertext Transfer Protocol); Server Log Files, Cookies; and Web Bugs.

The Internet & TCP/IP

The Internet, a network of interconnected computer networks, uses the TCP/IP Protocol for reliable delivery of packets of data that make up messages being passed from one software application another. Most often, the messages are a request from a client application for service from a service providing application.

Client / Server Computing and HTTP (Hypertext Transfer Protocol);

Using the Internet and TCP/IP the client application sends a request to a server application. The server sends a response to the client. This process requires that a connection is established between a computer hosting the client application and the computer hosting the server application. This connection is characterized as "stateless" because the clients as well as the server have no memory of prior connections. Furthermore, the server cannot distinguish one client request from another client.

Server Log Files

Each time a client requests a resource the server of that resource may record the following in its log files: the name & IP address of the client computer; the time of the request; the URL that was requested; the time it took to send the resource;; any

errors that occurred; the referer link and the kind of web browser that was used. If HTTP authentication was used, the username of the user of the client will be recorded.

Cookies

Cookies are used to solve the "Statelessness" of the HTTP Protocol and to store and retrieve user-specific information on the web. When an HTTP server responds to a request from a client it may send additional information that is stored on the client. This is "state information" or a "cookie". This cookie is used when client makes a request to the server that sent the cookie. The client will return the "cookie" that contains the client's state information. This state information may be a client ID that can be used as an index to a client data record on the server that provides the cookie.

Web Bugs

Web bugs are used to track a user's "clickstream" which a sequence of requests made by a specific client. The tutorial will provide an illustration of how web bugs function.

DATA COLLECTION IN WEB ANALYTICS

The above technology is used to collect data on web page usage. Data collection is carried out by using the clickstreams of a large number of clients. The data sources that make up these clickstreams are server log files and page tagging.

Server Log Files

Server log files use passive data collection. As the part of web browser/ web server transaction web servers provide resources to requesting clients. The web servers "log" these requests and a text file is written to memory recording attributes of these requests. As an example, consider the following HTTP request to a web server.

```
127.0.0.1 - frank [10/Oct/2000:13:55:36 -0700]
"GET /apache_pb.gif HTTP/1.0" 200 2326
```

The web server records the following data: 127.0.0.1 is the remote host (IP of the requesting computer); frank is the user name; [10/Oct/2000:13:55:36 -0700] is the date of the request; "GET /apache_pb.gif HTTP/1.0" is the request type; 200 is the status of the request and 2326 is number of bytes of the request.

Page Tagging

Page tagging is active data collection. It often requires a third party to implement. If page tagging is implemented by third party vendor they do the follwoing: supply page tags (web bugs); collects the data; analyzes the data. The party who contracted for the service accesses the data online or the vendor send reports sent to the party who contracted the service.

DATA ANALYSIS IN WEB ANALYTICS

The process for web data analysis requires three steps. The techniques used for Data Preparation, Pattern Discovery, and Pattern Analysis will be described and examples presented.

Data Preparation

Using data from the clickstream, data abstractions are performed concerning web usage, content, and structure. In particular these data abstractions establish precise semantics for the concepts: web site, users or visitors, user sessions, server sessions or visits, pageviews, and clickstreams. Further data abstractions concerning visitors establish precise semantics for the concepts: unique visitor, conversion rate, abandonment rate, attrition, loyalty, frequency, and recency.

Pattern Discovery

Statistical and data mining techniques are applied to the processed clickstream data in order to discover useful patterns.
The following data mining methods and algorithms have been adapted for the Web domain. These are: Association Rules; Sequential Pattern Discovery; Clustering and Classification.

Pattern Analysis

Patterns that have been discovered from the clickstream are subjected to further analysis. Common ways of analyzing such patterns are: using a query mechanism on a database where the results are stored; loading the results into a data cube and then performing OLAP operations. Visualization techniques are also used for an easier interpretation of the results.
These results in association with content and structure information concerning the Web site provide useful knowledge that can be to modify a web site according to the correlation between users and content.

TOOLS AND CASES STUDIES

To illustrate the use of Web Analytics the software tool provided by VisiStat - www.visistat.com - will be presented. Its use will be demonstrated in several case studies.

RESOURCES

1. Sterne J. (2002) Web Metrics: Proven Methods for Measuring Web Site Success, Wiley Publishing, Inc.
2. Davis, H. (2006) Google Advertising Tools, O'Reilly, Sebastopol, CA.
3. Peterson, E., (2005) Web Site Measurement Hacks, O'Reilly, Sebastopol, CA.
4. Tyler, M., & Ledford, J. (2006) Google Analytics, Wiley Publishing, Inc.
5. Peterson, E. (2004) Web Analytics Demystified, CafePress.
6. Turban, E., Aronson, J., Liang, T., Sharda, R. Decision Support and Business Intelligence Systems, 8/E, Prentice Hall
7. Visistat web site – http://www.visistat.com/ - Last accessed Feb. 12, 2008.